AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR018417

Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model

Manuela Girotto[1,2], Gabriëlle J. M. De Lannoy[1,2], Rolf H. Reichle[1], and Matthew Rodell[1]

[1]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, [2]Universities Space Research Association, Columbia, Maryland, USA

Key Points:
- A new scheme for GRACE-TWS assimilation is proposed
- The assimilation of GRACE-TWS primarily affects groundwater and has smaller impacts on soil moisture
- The assimilation of GRACE-TWS is affected by the use of observation scaling parameters

Correspondence to:
M. Girotto,
manuela.girotto@nasa.gov

Citation:
Girotto, M., G. J. M. De Lannoy, R. H. Reichle, and M. Rodell (2016), Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model, *Water Resour. Res.*, *52*, 4164–4183, doi:10.1002/2015WR018417.

Received 24 NOV 2015
Accepted 5 MAY 2016
Accepted article online 9 MAY 2016
Published online 27 MAY 2016

Abstract Observations of terrestrial water storage (TWS) from the Gravity Recovery and Climate Experiment (GRACE) satellite mission have a coarse resolution in time (monthly) and space (roughly 150,000 km^2 at midlatitudes) and vertically integrate all water storage components over land, including soil moisture and groundwater. Data assimilation can be used to horizontally downscale and vertically partition GRACE-TWS observations. This work proposes a variant of existing ensemble-based GRACE-TWS data assimilation schemes. The new algorithm differs in how the analysis increments are computed and applied. Existing schemes correlate the uncertainty in the modeled monthly TWS estimates with errors in the soil moisture profile state variables at a single instant in the month and then apply the increment either at the end of the month or gradually throughout the month. The proposed new scheme first computes increments for each day of the month and then applies the average of those increments at the beginning of the month. The new scheme therefore better reflects submonthly variations in TWS errors. The new and existing schemes are investigated here using gridded GRACE-TWS observations. The assimilation results are validated at the monthly time scale, using in situ measurements of groundwater depth and soil moisture across the U.S. The new assimilation scheme yields improved (although not in a statistically significant sense) skill metrics for groundwater compared to the open-loop (no assimilation) simulations and compared to the existing assimilation schemes. A smaller impact is seen for surface and root-zone soil moisture, which have a shorter memory and receive smaller increments from TWS assimilation than groundwater. These results motivate future efforts to combine GRACE-TWS observations with observations that are more sensitive to surface soil moisture, such as L-band brightness temperature observations from Soil Moisture Ocean Salinity (SMOS) or Soil Moisture Active Passive (SMAP). Finally, we demonstrate that the scaling parameters that are applied to the GRACE observations prior to assimilation should be consistent with the land surface model that is used within the assimilation system.

1. Introduction

Accurate profile soil moisture estimation is crucial to the quality of most water-related environmental, weather, and climate forecasts [*Koster et al.*, 2010]. Soil moisture controls the exchange of water and energy between the land surface and the atmosphere through evaporation and plant transpiration. Because soil moisture varies greatly in time and space, including in the vertical dimension, estimating profile soil moisture at regional to global scales remains a major challenge [*Hirschi et al.*, 2014; *De Lannoy and Reichle*, 2015].

Unlike microwave-based satellite missions that are sensitive only to surface soil moisture, the Gravity Recovery and Climate Experiment (GRACE) mission is unique because it provides highly accurate (~10–100 mm error) [*Wahr et al.*, 2006; *Swenson et al.*, 2006], column-integrated estimates of terrestrial water storage (TWS) variations (and its errors), after correcting for atmospheric and solid earth contributions. The TWS is the sum of groundwater, soil moisture, snow, surface water, ice, and biomass [*Tapley et al.*, 2004].

The TWS data are derived from highly precise, continuous measurements of the range (inter-satellite separation) and range-rate of GRACE's two coorbiting satellites [*Swenson and Wahr*, 2002]. Since its launch in March 2002, GRACE has provided unprecedented observations of water storage dynamics at basin to continental scales, which have improved the quantification and understanding of hydrologic states and fluxes [e.g., *Famiglietti and Rodell*, 2013]. For example, GRACE data have been valuable for drought characterization [*Houborg et al.*, 2012; *Thomas et al.*, 2014], identification and quantification of groundwater losses in the

world's major aquifer systems [*Rodell et al*., 2009; *Voss et al*., 2013], identification of regional flood potential [*Reager and Famiglietti*, 2009; *Reager et al*., 2014], quantification of snow cover and volume variations [*Frappart et al*., 2006; *Niu et al*., 2007], estimation of evapotranspiration in major river basins [*Rodell et al*., 2011], and quantification of ice mass loss over Antarctica, Greenland, and Alaskan glaciers [*Luthcke et al*., 2013; *Velicogna et al*., 2014].

The major limitations of GRACE-based TWS observations are related to their monthly temporal and coarse spatial resolution (roughly 150,000 km^2 at midlatitudes) [*Rowlands et al*., 2005; *Swenson et al*., 2006], and the vertical integration of the water storage components. These challenges can be addressed via data assimilation [*Zaitchik et al*., 2008]. Through the use of an appropriate observation operator [*Reichle et al*., 2014], assimilation techniques have the potential to (i) partition the vertically integrated GRACE-TWS observations into their surface and subsurface water components and (ii) downscale GRACE-TWS information to finer spatial and temporal scales.

The assimilation method employed by *Zaitchik et al*. [2008] and later by *Forman et al*. [2012], *Li et al*. [2012], *Houborg et al*. [2012], and *Li and Rodell* [2014] is similar to an ensemble smoother approach, a "two-step" scheme in which the land model integration is performed twice over the course of the same month: first to collect monthly TWS observation minus forecast differences, and a second time to update that month's simulated TWS. These early studies assimilated basin-averaged TWS observations using uniformly distributed observation errors (~20 mm). Subsequent research suggests that TWS assimilation at subbasin scales is preferable to assimilating basin-average observations [*Su et al*., 2010; *Forman and Reichle*, 2013; *Eicker et al*., 2014].

Other work suggested replacing the "two-step" assimilation scheme with a straight application of sequential Kalman filtering techniques [*Su et al*., 2010; *Eicker et al*., 2014; *Tangdamrongsub et al*., 2014] in which the increments are simply applied at the end of the assimilation window without rewinding the land surface model. Recently, *Eicker et al*. [2014], *Tangdamrongsub et al*. [2014], *van Dijk et al*. [2014], and *Kumar et al*. [2016] further explored GRACE data assimilation using $1^\circ \times 1^\circ$ gridded GRACE-TWS observations (rather than a basin or subbasin average estimates of TWS). *Kumar et al*. [2016] also investigated the use of the multiplicative gain factors [*Landerer and Swenson*, 2012] to restore signal amplitude that was dampened by processing of the GRACE gravity data into TWS anomaly fields.

The overarching objective of the present work is to determine whether soil moisture and groundwater estimation can be improved through the assimilation of GRACE-based TWS observations into a land surface model. To this end, we revisit various aspects of GRACE-TWS data assimilation systems, including (i) the computation and application of the data assimilation increments, given monthly coarse-scale TWS observation minus forecast residuals and (ii) the use of scaling parameters that are specifically derived to adjust the dynamic range of the observed TWS variations. We propose a revised assimilation scheme that computes increments that are less sensitive to the specific conditions on a single day within the month. All experiments conducted here use a three-dimensional (3-D), spatially distributed, ensemble-based approach to assimilate the gridded GRACE-TWS data product.

2. Data and Methods

2.1. Land Surface Model and Study Area

In line with previous work by *Zaitchik et al*. [2008], *Forman et al*. [2012], *Houborg et al*. [2012], *Li et al*. [2012], and *Li and Rodell* [2014], this study uses the catchment land surface model (CLSM) [*Koster et al*., 2000]. CLSM is the land model component of the Goddard Earth Observing System, version 5 (GEOS-5) modeling and data assimilation framework developed by the Global Modeling and Assimilation Office at the NASA Goddard Space Flight Center. CLSM differs from traditional, layer-based land surface models in that it includes an explicit treatment of the spatial variation of the soil water and water table depth within each hydrological catchment, as well as its effect on runoff and evaporation [*Koster et al*., 2000]. Subgrid hydrological processes are based on each catchment's topographical statistics, soil texture, and hydraulic parameters. CLSM's ability to represent shallow groundwater storage changes, which many global land surface models lack, is the main reason it has been targeted for GRACE-TWS data assimilation by this and previous studies.

Table 1. Ensemble Perturbation Parameters[a]

	Type	Standard Deviation	x,y_{corr}	t_{corr} (day)	Cross-Correlation With Perturbations in		
					pcp	sw	lw
pcp	M	0.5	2°	3	n/a	−0.8	0.5
sw	M	0.3	2°	3	−0.8	n/a	−0.5
lw	A	20 W m^{-2}	2°	3	0.5	−0.5	n/a
catdef	A	0.15 kg m^{-2} h^{-1}	2°	1			
srfexc	A	0.06 kg m^{-2} h^{-1}	2°	1			
swe	M	0.0012	2°	1			

[a]Multiplicative (M) or Additive (A) perturbations are applied to precipitation (pcp), incoming solar radiation (sw), incoming longwave radiation (lw), catchment deficit (*catdef*), surface excess (*srfexc*), and snow water equivalent (*swe*). Spatial correlations are indicated as x, y_{corr} and temporal correlations as t_{corr}.

CLSM does not model surface water hydrology (such as lakes and rivers). This represents a major limitation and a source of uncertainty in the modeled water storages in regions where surface water storage changes are a significant or dominant component of the terrestrial water storage signal, such as the wet tropics. For the United States, this applies to the immediate proximity of major rivers such as the Mississippi River, the Missouri River, and the Colorado River [*van Dijk et al*., 2014]. However, previous studies of TWS variability [e.g., *Rodell et al*., 2007] have noted that surface water occurs where the water table intersects the land surface, hence surface water and groundwater may be considered a single resource [*Winter et al*., 1998].

Snow water storage in CLSM is estimated by a three-layer snow model [*Stieglitz et al*., 2001]. The model defines three prognostic variables that describe the equilibrium soil moisture profile and deviations from the equilibrium across the entire grid cell (or computational unit), i.e., the catchment deficit (*catdef*), root-zone excess (*rzexc*), and surface excess (*srfexc*). The model prognostic catchment deficit (*catdef*) [*Ducharne et al*., 2000] is defined as the average depth of water that would need to be added in order to bring the catchment to saturation and is directly related to the unconfined mean groundwater table depth. Root-zone excess (*rzexc*) is defined as the amount of water in the root-zone layer (0–100 cm) in excess of the water that would be present if the entire soil moisture profile were in equilibrium. Surface excess (*srfexc*) is similarly defined for the surface layer (0–5 cm). Note that *rzexc* and *srfexc* may be positive or negative.

In extremely dry conditions, *catdef* approaches the volume of the dry pore space at the wilting point, which is controlled by, among other parameters, the depth-to-bedrock. *Houborg et al*. [2012] and *Li et al*. [2012] found that in some cases the CLSM model parameters do not permit a sufficiently large dynamic range to capture the extreme TWS anomalies observed during extended dry periods. As a work-around, these authors increased the depth-to-bedrock parameter. Here we avoid this complication by scaling the GRACE-TWS anomaly observations with scaling parameters that are consistent with the land model in our data assimilation system (see section 2.2 for details). This work uses a revised and improved treatment of soil texture (including organic matter) and associated soil hydraulic parameters for large-scale land surface models as described by *De Lannoy et al*. [2014].

The meteorological forcings used in Catchment are obtained from the Modern Era Retrospective Analysis for Research Application (MERRA) product [*Rienecker et al*., 2011].

The ensemble-based assimilation system estimates errors by applying perturbations to the land model prognostic and forcing variables and then diagnosing the ensemble spread. Specifically, we perturb select model prognostic variables related to soil moisture and snow mass (*catdef*, *srfexc*, and *swe*) and select surface meteorological forcing variables (precipitation and solar and longwave radiation). Twenty-four ensemble members were used to represent these errors. This ensemble size has been demonstrated to be suitable for GRACE data assimilation applications [e.g., *Zaitchik et al*., 2008; *Forman et al*., 2012]. Horizontal correlation lengths of the perturbations were chosen to be isotropic 2°, in order to represent the error scale of precipitation dynamics [*Reichle and Koster*, 2003]. The temporal correlation of the perturbations was chosen to be 3 days for the forcing fields, and 1 day for the prognostic states. Cross correlations were imposed between perturbations (i.e., errors) in solar radiation and precipitation (−0.8), solar and longwave radiation (−0.5), and precipitation and longwave radiation (0.5). The perturbation settings are summarized in Table 1 and are consistent with earlier studies [*Zaitchik et al*., 2008; *Forman et al*., 2012; *Houborg et al*., 2012].

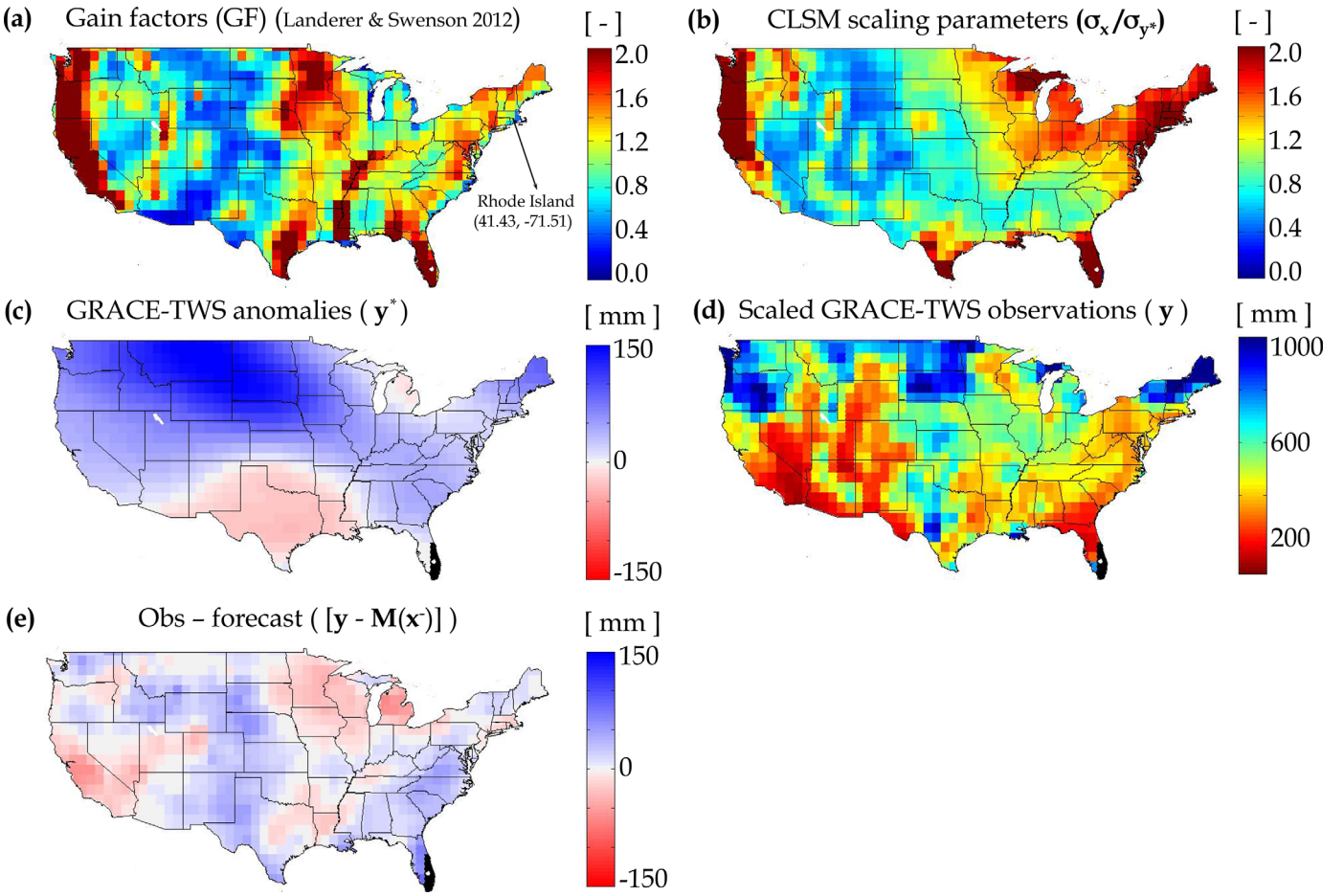

Figure 1. (a) Jet Propulsion Laboratory multiplicative gain factors (*GF*), [*Landerer and Swenson*, 2012]. (b) Ratio between CLSM versus GRACE variability (σ_x/σ_{y^*}, equation (2)). (c) Observed gridded (1° × 1°) GRACE-TWS anomalies (y^*), (d) scaled GRACE-TWS observations (y, equation (2)), and (e) difference between y (Figure 1c) and the corresponding forecast ($\mathbf{M}(\mathbf{x}^-)$, not shown) for February 2011.

The study domain is the contiguous United States (CONUS), and the experiment period covers 1 January 2003 to 1 January 2015. The model grid spatial resolution is 36 km on the equal area scalable earth (EASE version 2) grid [*Brodzik et al.*, 2012]. Model initial conditions were spun-up by looping the model twice through the 10 years from 1 January 1993 to 1 January 2003.

2.2. GRACE Terrestrial Water Storage Observations

GRACE observes temporal variations of the Earth's gravitational potential. The terrestrial water storage [*Landerer and Swenson*, 2012] data used in this study were obtained from the level-3 GRACE monthly 1° × 1° land gridded product available from the Jet Propulsion Laboratory (JPL; http://GRACE.jpl.nasa.gov). The data used in this work are a truncated (at spectral degree of 60) and smoothed (using a 300 km Gaussian filter) [*Landerer and Swenson*, 2012] version of the RL05 spherical harmonics from the Center for Space Research at the University of Texas. Spatial averaging, or smoothing, of GRACE data is necessary to reduce the contribution of noisy short wave-length components of the gravity field solutions [*Swenson and Wahr*, 2006], thus limiting random and systematic errors due to satellite and misrepresentation uncertainty [*Swenson and Wahr*, 2002]. However, this also implies that spatial scales finer than a few hundred kilometers are not resolved by GRACE observations [*Landerer and Swenson*, 2012], and along with the error reduction comes some loss of signal. As a means to restore the lost signal, JPL distributes multiplicative gain factors (*GF*, Figure 1a) obtained as in *Landerer and Swenson* [2012]. These gain factors are derived by mimicking the GRACE-TWS data filtering process on land surface simulations of TWS. Specifically, the gain factors

were obtained by minimizing the difference between "synthetic" monthly true (x_{true}) and filtered TWS simulations (x_{smooth}):

$$min\left\{\sum_{m=1}^{N_{months}}\left(x_{true,m}-GF\cdot x_{smooth,m}\right)^2\right\} \quad (1)$$

In their work, *Landerer and Swenson* [2012] used Global Land Data Assimilation System (GLDAS-NOAH) modeled TWS as x_{true}, and use a number of months (N_{months}) equal to 84 (i.e., from January 2003 to December 2009). While the purpose of these gain factors is primarily to correct for signal loss, there may be some sensitivity to the model used in deriving these factors [*Long et al.*, 2015]. Instead of applying the JPL gain factors (*GF*), here we downscale the GRACE-TWS observations via data assimilation (section 2.3.1).

Prior to data assimilation, we scale the observations to the long-term mean and variability of the model to avoid changing the model's climatology. Each monthly 1° × 1° GRACE-TWS observation represents, by design, the surface mass deviation (anomaly) for that month relative to the baseline average from 1 January 2004 to 31 December 2009. To obtain absolute observed TWS estimates (y) for the period 1 January 2003 to 1 January 2015, the observations (y^*) are scaled as follows:

$$y=(y^*-m_{y^*})\frac{\sigma_x}{\sigma_{y^*}}+m_x \quad (2)$$

where y^* are the truncated and smoothed GRACE-TWS anomalies, y are the adjusted (scaled) GRACE-TWS observations used in the data assimilation, and m_x and m_{y^*} are the 12 year averages of monthly simulated TWS (x) and GRACE-TWS anomalies (y^*), respectively. σ_x and σ_{y^*} are the corresponding long-term monthly standard deviations. In other words, the GRACE observations are scaled such that their long-term mean and standard deviation match those of the land surface model integrations. The modeled TWS statistics (m_x and σ_x) are obtained from model-based estimates of the GRACE-TWS observations, that is, from modeled TWS estimates at the spatial and temporal resolution of the GRACE observations (which are referred to as "observation predictions" in section 2.3.1).

Note that the a priori scaling approach does not imply that the climatology of the model is more correct than that of the observations. In fact, scaling GRACE observations to the modeled climatology is undesirable because it disregards potentially valuable information in the observations. Ideally, the model would be recalibrated to match the climatology of the GRACE observations, but this undertaking is not trivial, especially in the context of the operational GEOS-5 modeling system used here, and it is well beyond the scope of the present paper. For now, the scaling approach offers a feasible way of addressing the need for climatological consistency between observations and simulations in data assimilation systems [*De Lannoy et al.*, 2007; *Draper et al.*, 2015]. Revising the model to achieve such consistency is left for future work.

Figure 1b shows the scaling parameters σ_x/σ_{y^*}, i.e., the ratio between the standard deviations in TWS simulated by CLSM and observed by GRACE. The differences in the multiplicative parameters shown in Figures 1a and 1b are explained by the different design and purpose of these parameters. Parameters larger than one amplify the TWS observations. For the product-based gain factors (Figure 1a), this is, for example, the case along the West Coast and Florida. Parameters less than one instead reduce the amplitude of the observations (e.g., in the central U.S.). As an illustration, Figure 1c shows an example of truncated and smoothed GRACE-TWS anomalies (y^*) for February 2011. Most of the domain is wetter than the long-term average, which simply reflects the fact that winter-spring is a wet time of the year in the domain (in terms of TWS). The anomalously dry TWS conditions in Texas are a reflection of the severe drought that was ongoing in February 2011. Figure 1d shows the corresponding scaled GRACE-TWS observations (y) for the same month. The scaled observations now reflect primarily the long-term mean TWS conditions in the model. The assimilated information is the difference between the scaled observations of Figure 1d and the corresponding model forecast, that is, the observation minus forecast residuals or "innovation" (which is further discussed in the next section 2.3.1). This difference is shown for February 2011 in Figure 1e. For this specific month, the difference is positive, for example, in the Great Plains and Atlantic Coastal Plain regions, indicating that the model predicts TWS that is drier than the observed TWS, which will result in positive (wetting) increments to the model. Negative differences can be seen, for example, in California, Wisconsin, Minnesota, and part of Texas, where the assimilation will remove water from the model.

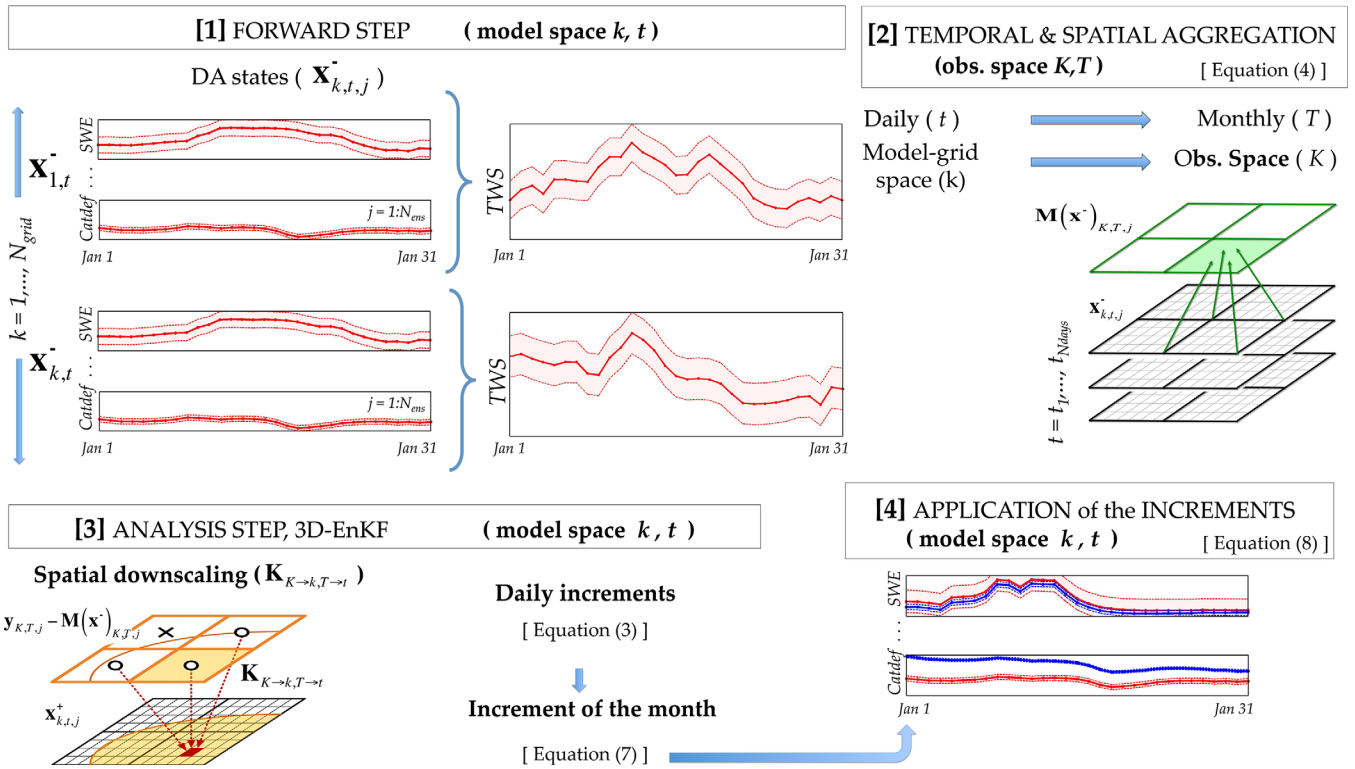

Figure 2. Simplified flowchart of the GRACE data assimilation (*DA*) system. [1] Conduct 1 month forecast ensemble integration without assimilation; store daily estimates of the *DA* state vectors ($\mathbf{x}^-_{k,t,j}$). [2] Calculate model terrestrial water storage (TWS) observation prediction through spatial aggregation from model to observation grid space and temporal aggregation from daily to monthly TWS estimates. [3] Calculate the increments via ensemble Kalman filter analysis. [4] Apply increments and integrate the model from the first day to the last day of the month and apply increments. At the end of the month, repeat from [1] for the next month. See section 2.3 for details.

2.3. Data Assimilation

A 3-D ensemble Kalman filter (EnKF) assimilation approach is used to merge monthly GRACE observations with model simulations. The "3-D" notation refers to the fact that the filter distributes information horizontally as well as vertically [*Reichle and Koster*, 2003; *De Lannoy et al.*, 2010]

Figure 2 illustrates the main steps of the GRACE-TWS assimilation system proposed in this paper: [1] the forward model is run for 1 month, during which state variables relevant to the assimilation scheme are stored in memory; [2] at the end of the forward model run, the monthly TWS observation predictions (or model forecasts) are calculated; [3] increments are calculated during the analysis step from the residuals between the (scaled) observations and the model forecasts; and, finally, [4] the dynamical model is rewound to the beginning of the month, the increments are applied to the model state variables, and the second 1 month forward model run is initialized and executed, which completes the cycle.

Two components of this algorithm are of particular interest in this paper: (i) the calculation of the instantaneous increments as representative of the monthly average increment (section 2.3.3), with an application of the increments as the initial water surplus or deficit (section 2.3.4); and (ii) the scaling of the gridded GRACE observations to ensure a climatologically consistent assimilation system (section 2.2, equation (2)).

2.3.1. The 3D-EnKF

In this study, multiple 1° × 1° gridded TWS observations around each 36 km model grid cell (k) are used to compute the increments ($\Delta\mathbf{x}_{k,t,j}$) for that grid cell as follows:

$$\Delta\mathbf{x}_{k,t,j} = \mathbf{K}_{K\to k,T\to t} \cdot [\mathbf{y}_{K,T,j} - \mathbf{M}(\mathbf{x}^-)_{K,T,j}] \tag{3}$$

where $\mathbf{K}_{K\to k,T\to t}$ is the (Kalman) gain matrix, $\mathbf{y}_{K,T,j}$ is the vector of observations, and $\mathbf{M}(\mathbf{x}^-)_{K,T,j}$ are the corresponding model predictions of the observations. $\mathbf{M}(\cdot)$ is also called the observation operator [*Reichle et al.*, 2014]. The subscript k refers to the 36 km model grid cell, while subscript K refers to the collection of

observations included in the update of grid cell k; t refers to the time for which the analysis increments are computed (see definition below); for example, $t = t_1, \ldots, t_{N_{days}}$, where N_{days} refers to the number of days within a month; T refers to the analysis time window across which the TWS observations and model forecasts are calculated (e.g., T = monthly); and j refers to the ensemble member ($j = 1, \ldots, N_{ens}$, where N_{ens} is the ensemble size). Note that the observations are perturbed as in *Burgers et al.* [1998] (see also section 2.3.2).

$\mathbf{M}(\mathbf{x}^-)_{K,T,j}$ is a vector of all 1° × 1° TWS observation predictions (i.e., modeled terrestrial water storage averaged to the temporal and spatial scale of the observations) within the time window (i.e., T = monthly) and within a 9° influence radius area (localization) around the grid cell (k) in question, with one element l given by (see step [2] of Figure 2):

$$[\mathbf{M}(\mathbf{x}^-)_{K,T,j}]_l = \frac{1}{N_{days}} \frac{1}{N_{grid,l}} \sum_{t=t_1}^{t_{N_{days}}} \sum_{k=1}^{N_{grid,l}} TWS_{k,t,j} \tag{4}$$

where N_{days} is the number of days in a month and $N_{grid,l}$ represents the number of 36 km grid cells within one GRACE observation (1°-radius). $TWS_{k,t,j}$ is one ensemble member (j) of the modeled terrestrial water storage for a given day (t) and 36 km model grid cell (k). Put differently, to obtain the model estimate of the 1° GRACE observation, we averaged TWS from all of the 36 km model grid cells whose center points are located within a circle with a 1° radius around the center point of the 1° gridded GRACE observation. This simplified approach to the resolution of the observations is supplemented by imposing a 3° horizontal error correlation scale for the observation errors (section 2.3.2).

The state vectors before the update ($\mathbf{x}^-_{k,t,j}$) and after the update ($\mathbf{x}^+_{k,t,j}$) are collections of CLSM prognostic variables (section 2.1) that make up TWS (instantaneous values at 0:00 UTC each day): *catdef*, *rzexc*, *srfexc*, canopy storage, and snow water equivalent (*swe*) [*Zaitchik et al.*, 2008; *Forman et al.*, 2012]. Details on the prognostic variables update are provided in section 2.3.4.

The gain ($\mathbf{K}_{K\to k,T\to t}$) is a "weighting matrix" that controls the amplitude of the update (increments) assigned to each variable of the state vector ($\mathbf{x}^-_{k,t,j}$), that is, the gain transforms observation-space TWS innovations $\left([\mathbf{y}_{K,T,j} - \mathbf{M}(\mathbf{x}^-)_{K,T,j}]\right)$ into model-space increments ($\Delta\mathbf{x}_{k,t,j}$). The weighting scheme is based (i) on the error cross correlations between each variable of the state vector (i.e., $\mathbf{x}^-_{k,t,j}$) and the terrestrial water storage observation prediction ($\mathbf{M}(\mathbf{x}^-)_{K,T,j}$) and (ii) on the uncertainties of the forecasts and the observations. Specifically, the gain matrix $\mathbf{K}_{K\to k,T\to t}$ is calculated as:

$$\mathbf{K}_{K\to k,T\to t} = [\mathbf{C}_{\mathbf{xM}}]_{K\to k,T\to t} \cdot \left([\mathbf{C}_{\mathbf{MM}}]_{K,T} + [\mathbf{C}_{\mathbf{yy}}]_{K,T}\right)^{-1} \tag{5}$$

where $[\mathbf{C}_{\mathbf{xM}}]_{K\to k,T\to t}$ is the error cross covariance between the state vector $\mathbf{x}^-_{k,t,j}$ and the observation prediction $\mathbf{M}(\mathbf{x}^-)_{K,T,j}$ (monthly averaged modeled TWS). The term $[\mathbf{C}_{\mathbf{xM}}]_{K\to k,T\to t}$ is particularly important because it determines the downscaling and vertical partitioning of the TWS innovations into fine-scale increments to individual water storage components (such as snow variables or soil moisture deficit/excess variables). Section 2.3.3 will further elaborate on $[\mathbf{C}_{\mathbf{xM}}]_{K\to k,T\to t}$. The term $[\mathbf{C}_{\mathbf{MM}}]_{K,T}$ is the error covariance of the monthly observation predictions and $[\mathbf{C}_{\mathbf{yy}}]_{K,T}$ is the observation error covariance (section 2.3.2).

2.3.2. Observation Error

Estimates of "total" GRACE-TWS uncertainties are needed to construct the observation error covariance $\mathbf{C}_{\mathbf{yy}}$ (equation (5)), which accounts for instrument errors and representativeness errors. Both components may be complicated functions of space and time. Instrument errors for GRACE-TWS depend on latitude and the smoothing radius of the spherical harmonics, with typical values for CONUS ranging from 15 to 30 mm [*Wahr et al.*, 2006]. Representativeness errors are associated with differences in resolution between the observations and model simulations [*Lahoz et al.*, 2010] as well as other discrepancies, such as leakage, or unmodeled processes, such as lake water storage, that are not resolved by the observation operator (equation (4)).

The observation error variances used here are derived by using a poor-man's adaptive filtering approach. Specifically, the diagonal elements (subscript "*dd*") of the covariance of the TWS innovations ($\mathbf{C}[\mathbf{y} - \mathbf{M}(\mathbf{x}^-)]_{dd}$) are given by the sum of diagonal elements of the observation error covariance ($[\mathbf{C}_{\mathbf{yy}}]_{dd}$) and the (observations-space) model forecast error covariance ($[\mathbf{C}_{\mathbf{MM}}]_{dd}$) [*Desroziers et al.*, 2005]:

$$\mathbf{C}[\mathbf{y} - \mathbf{M}(\mathbf{x}^-)]_{dd} = [\mathbf{C}_{\mathbf{yy}}]_{dd} + [\mathbf{C}_{\mathbf{MM}}]_{dd} \tag{6}$$

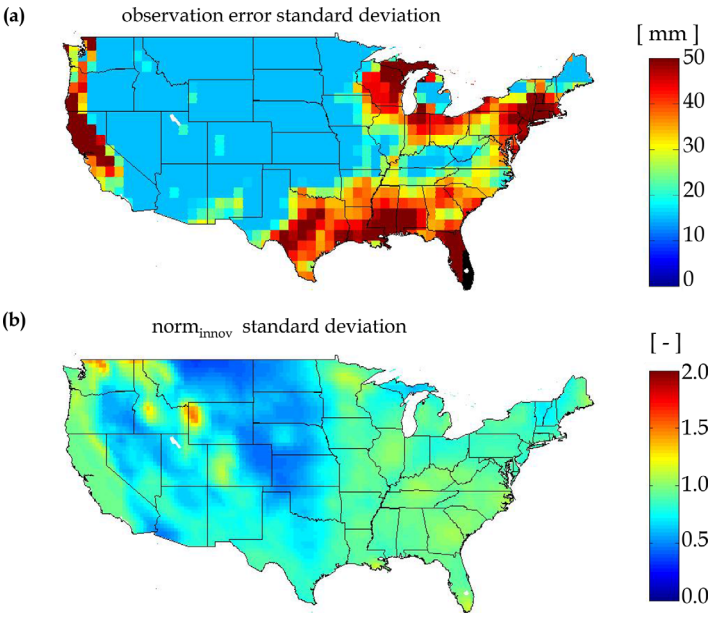

Figure 3. (a) Estimated TWS observation error standard deviation used in the data assimilation experiments (section 2.3.2) and (b) standard deviation of the normalized innovations ($norm_{innov}$) across all of the experiment period (1 January 2003 to 1 January 2015).

where the subscripts K and T have been omitted for clarity. To back out the observation error variance terms ($[\mathbf{C_{yy}}]_{dd}$), we estimated the other two terms from the innovations time series of an open-loop (no assimilation) experiment by substituting ensemble statistics with time series statistics (ergodicity principle). We imposed a minimum threshold value of 15^2 mm^2 for the observation error variance based on the results by *Wahr et al.* [2006]. This is by no means a perfect scheme and will lead to an imperfect, but to our best knowledge, reasonable estimate of the observation error to be used within the assimilation scheme.

Figure 3a shows the spatial distribution of the errors obtained from the poor-man's adaptive approach. The domain average of the observation error standard deviation is 22 mm. The highest observation errors are found in coastal regions, and a minimum of 15 mm observation error is imposed to the central areas of the CONUS domain (where the poor-man's adaptive approach would have set a smaller error).

One way to verify the optimality of the update step is by looking at the distribution of the monthly coarse-scale normalized innovations (i.e., $norm_{innov} = [\mathbf{y} - \mathbf{M}(\mathbf{x}^-)](\mathbf{C_{MM}} + \mathbf{C_{yy}})^{-1/2}$, omitting the subscripts K and T for simplicity). Figure 3b shows the temporal standard deviation of $norm_{innov}$. Values close to 1.0 indicate that the sum of the modeled forecast and observation error variances is close to the total variance of the actual errors as estimated from the innovations time series. The standard deviation of the $norm_{innov}$ is close to unity for the entire domain, except for the Great Plains region, where this metric is smaller than one and the actual observation and/or model forecast errors are overestimated.

Previous work addressed the fact that GRACE-derived TWS errors are highly correlated in space [*Forman and Reichle*, 2013; *Eicker et al.*, 2014]. In particular, these studies found that data assimilation of GRACE-TWS performed optimally when the spatial averaging scale was chosen to be on the order of 5° × 5°, a scale at which the observation errors become uncorrelated. Here we use a spatial correlation length of 3° for the observation errors to account for the fact that the errors in the 1° gridded GRACE-TWS observations are highly correlated.

2.3.3. Calculation of the Analysis Increments

Technically, a valid increment (equation (3)) can be calculated (and applied) for any time (t) within the observation and assimilation window (T). In earlier studies, the increments are typically calculated for a single instant in the month, either at the beginning or at the end of the month (section 1). For the new data assimilation scheme (DA) proposed here, we exploit the fact that depending on the choice of t within the month, the values of the increments vary in response to the changing relationship between errors in the state ($\mathbf{x}^-_{k,t,j}$) at a particular time t and errors in the (monthly) observation predictions ($\mathbf{M}(\mathbf{x}^-)_{K,T,j}$), because $[\mathbf{C_{xM}}]_{K\rightarrow k,T\rightarrow t}$ depends on the instantaneous structure of the model ensemble. To obtain increments that are representative of the entire month, we calculate increments $\Delta\mathbf{x}_{k,t,j}$ (equation (3)) for 00:00 UTC of each day of the month (i.e., $t = t_1, \ldots, t_{N_{days}}$). Then, the monthly average of these increments $\Delta\mathbf{x}_{k,T,j}$ is obtained as:

$$\Delta\mathbf{x}_{k,T,j} = \frac{1}{N_{days}} \sum_{t=t_1}^{t_{N_{days}}} \Delta\mathbf{x}_{k,t,j} \tag{7}$$

The single monthly averaged increment $\Delta\mathbf{x}_{k,T,j}$ is finally applied back to the model state at the beginning of the month (section 2.3.4).

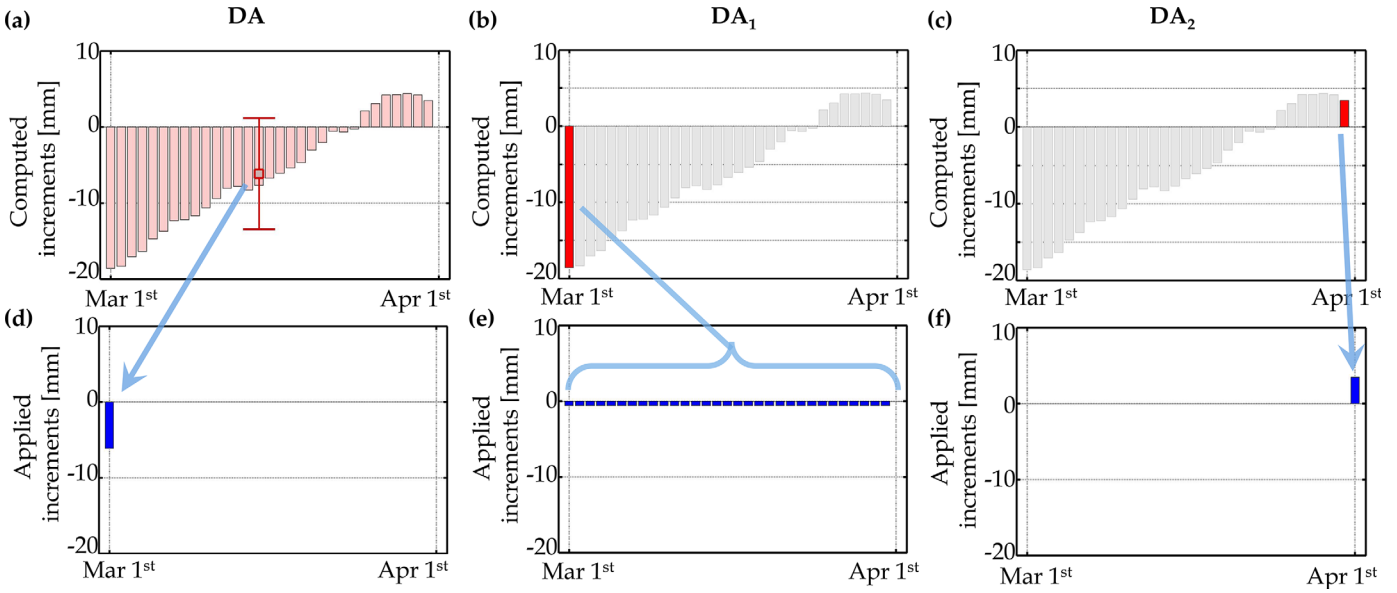

Figure 4. (a–c) Calculation and (d–f) application of the increments for (a, d) the newly proposed data assimilation approach *DA*, (b, e) the approach DA_1 [e.g., *Zaitchik et al*., 2008], and (c, f) the approach DA_2 [e.g., *Su et al*., 2010]. For a given algorithm, only the increments shown in red are computed. Increments shown in grey color in Figures 4b and 4c are not computed in approaches DA_1 and DA_2 and shown only for reference. See sections 2.3.3 and 2.3.4 for details.

The red bars in Figure 4a show an example of daily instantaneously calculated increments for one model grid cell, 1 month, and one variable (*catdef*) of the state vector. The monthly averaged increment (equation (7)) is shown by the red square, with the error-bar representing the monthly standard deviation. A similar approach was introduced by *Eicker et al*. [2014] where the monthly averaged increment was computed from an ensemble composed of temporally averaged water storage compartments. Both of these approaches differ from previous studies [*Zaitchik et al*., 2008; *Forman et al*., 2012; *Li et al*., 2012; *Houborg et al*., 2012; *Li and Rodell*, 2014; *Kumar et al*., 2016] that use a two-step assimilation approach but calculate increments for the first day of the month only (i.e., for $t = t_1$, as shown in Figure 4b). That is, in the previous studies, the gain and increments calculation relies solely on the cross-covariance $[\mathbf{C_{xM}}]_{K\rightarrow k,T\rightarrow t}$ between the monthly average TWS observation and the model state on the first day of the month. This approach is hereafter referred as "DA_1." Another alternative [*Tangdamrongsub et al*., 2014; *Su et al*., 2010] is to calculate increments for the end of the assimilation window (for $t{=}t_{N_{days}}$) as shown in Figure 4c. That is, increments are only sensitive to $[\mathbf{C_{xM}}]_{K\rightarrow k,T\rightarrow t}$ on the last day of the month. This latter option will be referred as "DA_2" (Table 2).

2.3.4. Application of the Analysis Increments

The monthly increments can be applied in various ways. In our assimilation scheme "*DA*," the monthly average increment (equation (7), Figure 4a) is applied in full at 00:00 UTC of the first day of the month ($t = t_1$, as illustrated in Figure 4d).

Specifically, the updated state vector ($\mathbf{x}^-_{k,t,j}$) becomes:

$$\mathbf{x}^+_{k,t=t_1,j}=\mathbf{x}^-_{k,t=t_1,j}+\Delta\mathbf{x}_{k,T,j} \tag{8}$$

where only *catdef*, *rzexc*, and *swe* are updated explicitly. The *srfexc* and canopy interception reservoir variables are not updated because they represent only a very small fraction of the TWS variations and are highly variable in space and in time. Any increments to these variables would inevitably be spurious. Note that we nevertheless include *srfexc* and canopy interception reservoir in the calculation of TWS (i.e., in the calculation of $\mathbf{M}(\mathbf{x})_{K,T,j}$). The application of the increments in *swe* is supplemented with an adjustment of snow heat content, snow cover extent, and snow depth. These adjustments assume that the modeled snow densities and temperatures remain unchanged during the analysis update. Note that the assimilation algorithm requires that the observation errors are uncorrelated in time. If we calculated and applied increments for each day of the month, we would repeatedly assimilate essentially the same observations, with nearly perfectly correlated errors. Therefore, we apply only a single mean increment for a given month.

Table 2. Open-Loop (*OL*) and Data Assimilation Experiment (*DA*, DA_1, and DA_2) Configurations

CASE ID	Update Type	Calculation of the Increments	Application of the Increments	Example
OL	none			
DA	Two step	Monthly average	$t = t_1$	This paper.
DA_1	Two step	$t = t_1$	Divide increment by $1/N_{days}$ and apply resulting fraction at $t = t_1, \ldots, t_{N_{days}}$	*Zaitchik et al.* [2008], *Forman et al.* [2012], *Houborg et al.* [2012], *Li et al.* [2012], *Li and Rodell* [2014], and *Kumar et al.* [2016]
DA_2	Sequential	$t = t_{N_{days}}$	$t = t_{N_{days}}$	*Su et al.* [2010] and *Tangdamrongsub et al.* [2014]

Our approach differs from the approach DA_1 of earlier GRACE data assimilation studies [*Zaitchik et al.*, 2008; *Forman et al.*, 2012; *Li et al.*, 2012; *Houborg et al.*, 2012; *Kumar et al.*, 2016] where the increment was calculated for the beginning of the month, then divided by the number of days in the month (N_{days}), and finally applied uniformly on each day within the month as illustrated in Figure 4e. In the following, this approach is referred to as DA_1. Our approach also differs from the sequential filtering approach DA_2 of *Su et al.* [2010] and *Tangdamrongsub et al.* [2014] where the increment was calculated for the end of the assimilation window and then applied in full at the beginning of the next assimilation window, as illustrated in Figure 4e. Table 2 summarizes the various approaches.

2.4. In Situ Observations and Metrics For Validation

Results obtained after the assimilation of GRACE-TWS observations are evaluated against independent in situ observations of groundwater (section 2.4.1) and soil moisture (section 2.4.2). All available soil moisture and groundwater measurements within the experiment period are used for validation at sites that provided at least 20 months of measurements. Monthly averages are computed only if at least 66% of the daily observations in a month are available.

2.4.1. Groundwater Observations

Groundwater observations were obtained from 348 monitoring wells maintained by the U.S. Geological Survey (USGS), and from 17 sites in the Shallow Groundwater Wells Network maintained by the Illinois State Water Survey (http://www.isws.illinois.edu/warm). Measurements are reported as depth-to-water-table from the land surface. We selected only measurements from wells in unconfined or semiconfined aquifers, because changes in confined aquifer head are not directly proportional to the mass changes observed by GRACE, nor does CLSM simulate confined aquifer storage. An aquifer was determined to be unconfined, semiconfined, or confined based on the metadata information available for that site, a review of literature describing the aquifer, and visual inspection of the magnitude and seasonality of the water depth variations. The selected wells typically display a clear seasonal cycle of water depth variations that are neither very large nor very small, and they lack sudden drops in the water table that may be associated with pumping. The quality control screening reduced the number of wells to a total of 181 that were deemed to have sufficient data of acceptable quality. Specific yield is used to convert the depth-to-water to water-equivalent-depth. We used specific yield values as derived by *Rodell et al.* [2007], *Houborg et al.* [2012], and *Li and Rodell* [2014]. The water-equivalent-depth observations (mm) are compared directly to the CLSM water deficit (*catdef*, section 2.1).

2.4.2. Root-Zone and Surface Soil Moisture Observations

Two sets of in situ root-zone and surface soil moisture measurements were compared to model soil moisture in the 0–100 cm "root-zone" layer and the 0–5 cm surface layer, respectively. The first set of soil moisture measurements is referred to as "Cal/Val" measurements. These are grid-cell-scale (36×36 km^2) averaged measurements, collected by the U.S. Department of Agriculture in experimental watersheds across the U.S. The surface measurements of this data set were originally obtained for the purposes of calibrating and validating remote sensing observations [*Jackson et al.*, 2012; *De Lannoy et al.*, 2014; *Entekhabi*, 2014]. We identified four watersheds with sufficient monthly observations of surface soil moisture: Reynolds Creek, ID, Walnut Gulch, AZ, Little Washita, OK, and Little River, GA [*Cosh et al.*, 2008; *Jackson et al.*, 2010; *Entekhabi*, 2014]. The last two watersheds also have root-zone measurements available for validation of the data assimilation results. These sites are unique because they provide spatially averaged soil moisture

measurements, and therefore they are particularly appropriate for validation of gridded estimates from land surface modeling and data assimilation.

Sparse networks provide more localized in situ soil moisture measurements that are generally difficult to compare directly to a model product [*Koster et al*., 2009]. Nonetheless, given the geographical extent of these measurements, and given that many of these sites provide information on root-zone soil moisture, sparse networks play an important role in evaluating model soil moisture. Data were obtained from two networks over the U.S., the Soil Climate Analysis Network (SCAN) [*Schaefer et al*., 2007] and the U.S. Climate Reference Network (USCRN) [*Diamond et al*., 2013]. Root-zone soil moisture estimates are calculated based on vertically weighted averages of measurements at 5, 10, 20, and 50 cm depth. After quality control [*Liu et al*., 2011; *De Lannoy et al*., 2014], we used 56 SCAN sites and 33 USCRN sites for the validation of surface soil moisture, and we used 53 SCAN sites and 30 USCRN sites for the validation of root-zone soil moisture. The number of sites used here is smaller than that of soil moisture assimilation studies [e.g., *Liu et al*., 2011; *De Lannoy et al*., 2014], because here the validation is conducted at the monthly scale and a minimum number of 20 monthly data pairs is required.

2.4.3. Validation Approach

The validation is performed against monthly averaged time series of in situ measurements, to match the temporal resolution of the GRACE-TWS observations. The statistical skill metrics include the correlation coefficient (R) and the unbiased root-mean-square difference (ubRMSD) [*Entekhabi et al*., 2010]. Note that we choose to refer to "differences" rather than "errors," because the in situ observations are not perfect and could also contain errors. The ubRMSD is computed as the RMSD after removing the long-term mean difference and is also known as the standard deviation of the differences. The R and ubRMSD metrics are commonly used to evaluate the mismatch between observations and data assimilation results in terms of dynamic variability (unitless R) and overall closeness (ubRMSD, with units of the evaluated variable) [*Entekhabi et al*., 2010].

For each site individually, the skill metrics and the 95% confidence intervals take into account the temporal autocorrelation of the monthly time series. The individual sites are then grouped spatially into clusters, and the metrics (and confidence intervals) for each cluster are computed by averaging across all sites within each cluster (that is, we do not assume that the metrics for two sites within the same cluster are independent). Finally, network average metrics are computed by averaging across clusters, where the average *CI* is further divided by the square root of the number of clusters, assuming that each cluster adds independent data for validation [*De Lannoy and Reichle*, 2015]. The clustering approach ensures that skill metrics from in situ sites in more densely sampled regions do not dominate the CONUS-average skill metric. The cluster-based averaging thus provides meaningful statistics and confidence intervals and enables us to determine the statistical significance of differences in skill between the experiments with and without GRACE data assimilation.

3. Results and Discussion

Sections 3.1 and 3.2 only discuss the results obtained with the newly proposed GRACE-TWS data assimilation scheme. Section 3.3 compares the new system with data assimilation schemes that differ in the computation and application of the increments, and in the treatment of the GRACE-TWS observations.

3.1. Comparison to Independent Soil Moisture and Groundwater Observations

Figure 5 shows the difference in skill (before clustering) between *DA* and the model-only, ensemble open-loop (*OL*) estimates, in terms of ubRMSD and R versus in situ measurements of soil moisture and groundwater at the individual sites. Table 3 reports the ubRMSD and R metrics calculated across the observed sites as described in section 2.4.3.

The *DA* generally improves groundwater estimates over the model-only open-loop simulations (Figures 5a and 5b) for the majority of the in situ locations. Improvements are particularly noticeable in the Mid-West, the Mississippi River basin, and the Atlantic Coastal Plain regions. For these regions, high groundwater depletion rates occurred during the period 2000–2008, possibly as a result of a temporary natural decrease in precipitation (interannual variability) or due to increased groundwater withdrawals [*Konikow*, 2013]. It is possible that the modeling system cannot simulate this, whereas GRACE-TWS detects the depletion, and

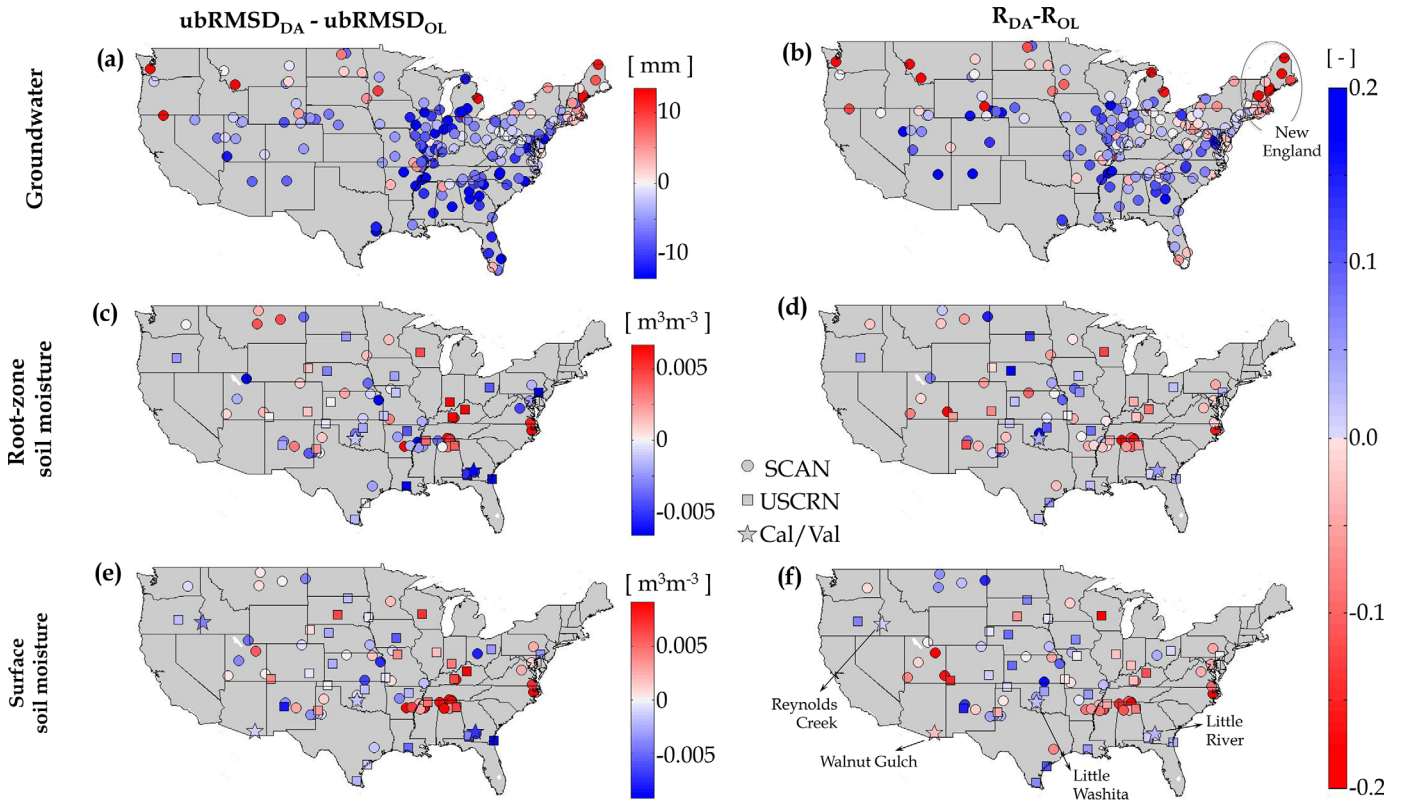

Figure 5. Difference in skill between the data assimilation (*DA*) and open-loop (*OL*; no assimilation) estimates for (a and b) groundwater, (c and d) root-zone soil moisture, and (e and f) surface soil moisture. Skill is measured as the (a, c, e) unbiased root-mean-squared difference (ubRMSD) and (b, d, f) correlation coefficient (R) versus in situ measurements. Blue colors indicate skill improvement, that is, *DA* is more skillful than *OL*, and red colors indicate skill degradation.

the data assimilation manages to correct what is otherwise unpredicted by the model. For some regions (including the West Coast, Montana, and New England), the *DA* estimates have degraded groundwater correlation skills compared to the *OL*. In these regions, either the modeled TWS or GRACE-TWS seasonality is out-of-phase with the seasonality indicated by the in situ groundwater observations, and GRACE-TWS assimilation cannot bring the results closer to in situ observations.

The inconsistency between modeled and observed seasonality could be due to a shortcoming of the model, such as its highly simplistic representation of aquifer recharge and storage, or its inability to represent water management. The largest inconsistencies occur in New England, where simulated subsurface water storage peaks in March or April. The in situ groundwater observations indicate maximum groundwater storage typically 1 month later, with a secondary maximum in December or January. This bimodal seasonality of groundwater can be explained as follows. There is a net increase in TWS from September to March (as observed by GRACE), during which period precipitation exceeds the sum of runoff and evapotranspiration. In New England, a significant portion of the winter precipitation occurs as snowfall, which accumulates on the (frozen) surface, reducing recharge during January and February. Subsequent snowmelt produces a large spike in recharge and maximum groundwater storage in April or May. While CLSM properly simulates snow accumulation in January and February, it fails to represent the winter recharge variability and delayed peak in groundwater storage.

Nonetheless, on balance the groundwater skill is improved with higher R values for *DA* at 114 sites, and lower ubRMSD values for *DA* at 140 sites out of the 181 validation sites (Table 3). The improvements are not, however, statistically significant, because of the limited sample size (monthly data): average ubRMSD values for groundwater are 64 $\pm$ 5 and 60 $\pm$ 5 mm for the *OL* and the *DA*, respectively; and average R values for groundwater are 0.58 $\pm$ 0.03 and 0.60 $\pm$ 0.03 for the *OL* and the *DA*, respectively (Table 3).

Mixed results are obtained for root-zone and surface soil moisture (Figures 5c–5f). Overall, the *DA* skill for soil moisture does not differ, in a statistical sense, from that of *OL* for all of the in situ observation types. At

Table 3. Mean, Median (Q50), and Interquantile Range (Q25, Q75) of the Correlation Coefficient (R) and the Unbiased Root-Mean-Square Difference (ubRMSD) Across All Validation Locations for Estimates From the Open-Loop (*OL*) and the Data Assimilation Scheme *DA*[a]

				R		ubRMSD		
		N. Sites		Mean (*CI*)	Q50 (Q25, Q75)	Mean (*CI*)	Q50 (Q25, Q75)	
TWS		CONUS	*OL*	0.69 (<±0.01)	0.70 (0.61, 0.79)	56 (<±0.01)	53 (41, 73)	mm
			DA	0.91 (<±0.01)	0.93 (0.89, 0.96)	28 (<±0.01)	25 (21, 37)	mm
GW		181	*OL*	0.58 (±0.03)	0.60 (0.51, 0.71)	64 (±5)	62 (51, 75)	mm
			DA	0.60 (±0.03)	0.64 (0.53, 0.74)	60 (±5)	56 (45, 69)	mm
			% improved	63%		77%		
Root zone	Cal/Val	2	*OL*	0.59 (±0.15)	n/a	0.028 (±0.011)	n/a	m^3m^{-3}
			DA	0.65 (±0.13)	n/a	0.024 (±0.010)	n/a	m^3m^{-3}
			% improved	100%		100%		
	USCRN and SCAN	83	*OL*	0.68 (±0.03)	0.71 (0.62, 0.78)	0.039 (±0.004)	0.038 (0.030, 0.048)	m^3m^{-3}
			DA	0.68 (±0.03)	0.69 (0.60, 0.80)	0.038 (±0.004)	0.039 (0.027, 0.049)	m^3m^{-3}
			% improved	43%		58%		
Surface	Cal/Val	4	*OL*	0.62 (±0.09)	n/a	0.032 (±0.010)	n/a	m^3m^{-3}
			DA	0.64 (±0.09)	n/a	0.029 (±0.009)	n/a	m^3m^{-3}
			% improved	75%		100%		
	USCRN and SCAN	89	*OL*	0.66 (±0.03)	0.71 (0.62, 0.74)	0.048 (±0.004)	0.048 (0.037, 0.058)	m^3m^{-3}
			DA	0.66 (±0.03)	0.69 (0.60, 0.78)	0.049 (±0.005)	0.047 (0.035, 0.063)	m^3m^{-3}
			% improved	49%		51%		

[a]Q50, Q25, and Q75 are not reported for the Cal/Val sites because only four sites are available. Mean statistics and 95% confidence intervals (*CI*) are obtained from clustering of the sites (section 2.4). Soil moisture metrics are computed by merging sites from the Soil Climate Analysis Network (SCAN) and the U.S. Climate Reference Network (USCRN). TWS metrics are computed against the assimilated (scaled) GRACE-TWS observations.

the Cal/Val sites, root-zone soil moisture correlation (R) values are 0.59 ± 0.15 and 0.65 ± 0.13 for the *OL* and *DA* case, respectively, indicating a small improvement from GRACE-TWS assimilation. Similarly, the ubRMSD decreases from 0.028 ± 0.11 mm for *OL* to 0.024 ± 0.010 mm for *DA* (Table 3). The *DA* case improves root-zone ubRMSD and R statistics for all of the Cal/Val watersheds (Table 3). Similar results can be seen in the statistics obtained for surface soil moisture, where three out of four Cal/Val surface soil moisture sites show improved correlation statistic skills with *DA*, and all of them exhibit improved ubRMSD. At the sparse network sites (SCAN and USCRN), however, the root-zone soil moisture ubRMSD is improved at only 58% of the sites, and R is degraded at 57% of the sites. Similarly, for surface soil moisture, the ubRMSD is improved at 51% of the sites and R is degraded at 51% of the sparse network sites. For example, soil moisture skill values are typically degraded along the northern edge of Alabama.

These results demonstrate that GRACE-TWS assimilation is somewhat more valuable for groundwater, and not yet sufficient to unambiguously improve the estimation of surface and root-zone soil moisture. It is not surprising for GRACE-TWS to have a smaller impact on surface soil moisture. In fact, the memory of surface and root-zone soil moisture is expected to be smaller than that of groundwater. Furthermore, the relative contribution of soil moisture to TWS is expected to be smaller than that of groundwater. The next section further discusses the impact of GRACE-TWS on the various vertical water storages of the soil moisture profile.

3.2. Downscaling of GRACE-TWS Observations

Data assimilation is a means to downscale the column integrated, spatially and temporally coarse GRACE-TWS observations. The methods described in section 2.3 translate information from the observation space to the model space by partitioning the differences between the observed and simulated TWS into increments to each modeled TWS component (that is, groundwater, soil moisture, snow, etc.) at finer spatial and temporal scales. The spatial and temporal patterns and the disaggregation of the increments into the TWS components are discussed in this section.

Figure 6 shows time series (for 1 year, 2008) of ensemble average instantaneous increments calculated at 00:00 UTC each day ($\Delta\mathbf{x}_{k,t}$) at two locations and for three of the assimilation state variables, surface excess (*srfexc*), root-zone excess (*rzexc*), and catchment deficit (*catdef*), which are the model prognostic variables used to diagnose soil moisture profile increments (section 2.1). The two locations are marked on the maps in Figure 7, and they correspond to a location in California's Central Valley (Figure 6a), and to the Little Washita Cal/Val site in Oklahoma (Figure 6b). Little Washita is fairly representative of the domain in terms of the calculation of the increments, while the Central Valley location stands out in that regard, as discussed

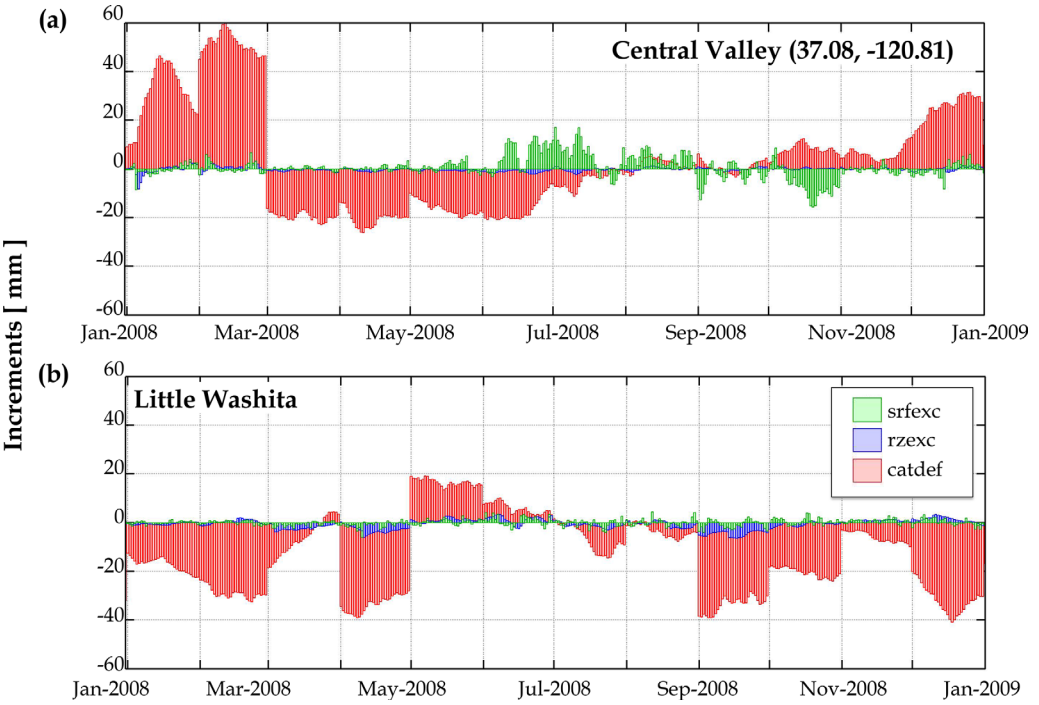

Figure 6. January 2008 to December 2008 ensemble average of daily instantaneous increments ($\Delta\mathbf{x}_{k,t}$) for the data assimilation state variables *srfexc* (green), *rzexc* (blue), and *catdef* (red) at (a) a location in California's Central Valley and (b) at the Little Washita "Cal/Val" site (see Figure 7 for locations). Note that within the *DA* scheme (Figure 2), the daily increments shown here are averaged into monthly mean values before they are applied to the model forecast.

below. At both of these locations, snow and canopy water storage are insignificant, and thus not discussed. For both locations, the total soil moisture profile increment time series is dominated by *catdef* increments (i.e., shallow groundwater), which range approximately between −60 and 60 mm, whereas the increments in the surface and root-zone soil moisture model prognostic variables (*rzexc* and *srfexc*) are 1 order of magnitude smaller. Thus, GRACE-TWS assimilation primarily affects (in absolute terms) *catdef*, which is associated with moisture over the entire profile depth and thus governs the groundwater estimates from the model.

Figure 6 also shows the day-to-day variability in increments within each month. The variability within the month is largest for the model prognostics associated with near-surface nonequilibrium soil moisture conditions (i.e., *srfexc* and *rzexc*), and lowest for the model prognostic variable associated with equilibrium moisture profile conditions (i.e., *catdef*). Increments can potentially change signs (i.e., surplus versus deficit of water) within the course of a single month. This situation occurs very frequently for *srfexc* and occasionally for *rzexc* (e.g., January 2008 for the Central Valley and June 2008 for Little Washita). The change of sign in the increments is less frequent but still possible for *catdef* (e.g., September 2008 for the Central Valley and March 2008 for Little Washita).

A particular case is the Central Valley location (Figure 6a) where the surface excess (*srfexc*) increments are small during the winter months, but become large during summer. This may be a result of the complicated hydrology that characterizes the Central Valley region with intensive water management and agricultural practices, which are not modeled in CLSM. Irrigation in particular would be consistent with the generally positive *srfexc* increments during summer. By assimilating GRACE-TWS, features that are missing in the model may be corrected.

Figure 7 shows maps of increment statistics for the assimilation state variables *srfexc*, *rzexc*, and *catdef*. The left column of Figure 7 shows the typical magnitude of the monthly mean increments, computed as the 12 year average (1 January 2003 to 1 January 2015) of the absolute values of the monthly mean, ensemble average increments ($\Delta\mathbf{x}_{k,T}$, equation (7)). The spatial mean values (spatial standard deviations) of the typical increments are 0.63 (0.39) mm for *srfexc*, 0.54 (0.25) mm for *rzexc*, and 15.29 (5.06) mm for *catdef*. This result

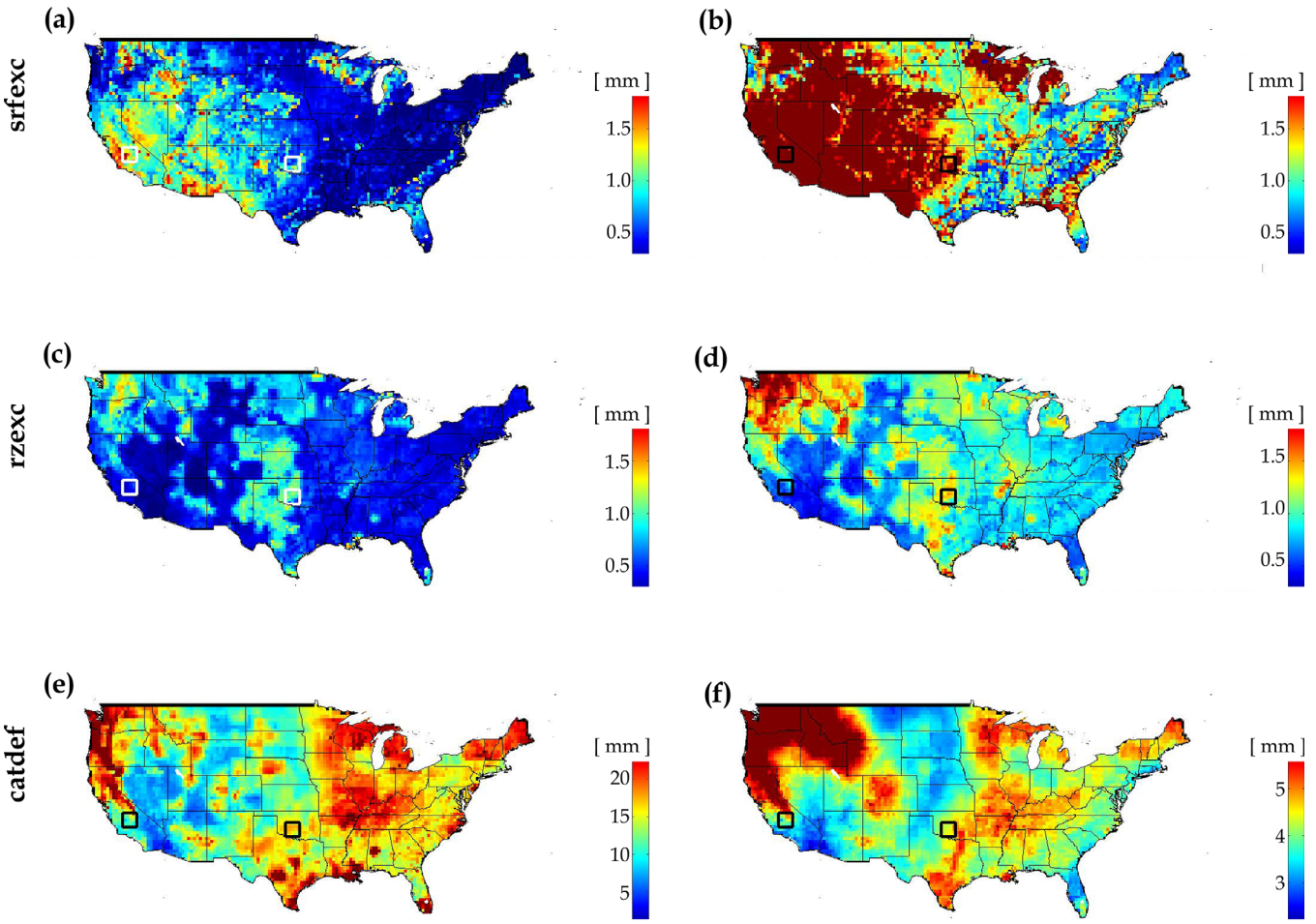

Figure 7. First January 2003 to first January 2015 average of the (a, c, e) typical (absolute) magnitude of monthly average increments and (b, d, f) average standard deviation of daily increments within a month in (a and b) *srfexc*, (c and d) *rzexc*, and (e and f) *catdef*. Squares indicate the location of the example time series shown in Figure 6.

confirms that GRACE-TWS assimilation has the biggest impact on the *catdef*, and therefore on groundwater. Since the GRACE observations were scaled prior to data assimilation (section 2.2), the increments only adjust for the timing of the water storage signals, or for processes that are not modeled (such as trends due to extensive groundwater depletion for irrigation purposes), but not for errors in the mean water storage or its variability.

The right column of Figure 7 shows the typical variability of the daily increments within each month (or intramonthly variability), computed as the 12 year average (1 January 2003 to 1 January 2015) of the standard-deviation of the ensemble average daily increments in each month. When the daily increments within a month vary a lot (i.e., high intramonthly variability), then using just the increments of the first day (as in DA_1), the last day (as in DA_2), or, for that matter, any single day within the month will likely result in suboptimal increments. Spatial mean values of the intramonthly standard deviation in the increments are 1.81 mm for *srfexc*, 0.95 mm for *rzexc*, and 4.49 mm for *catdef*. For *srfexc*, the largest values are found in the Western (driest) regions of CONUS. For *rzexc*, the intramonthly variability of the increments is greatest in the Northwest and the Great Plains region. Similarly, *catdef* increments tend to have the greatest intramonthly variability in the Northwest. These areas of large intramonthly variability are roughly collocated with areas where the typical (absolute) increments are largest (left column of Figure 7). The intramonthly variability in the increments for the surface and root-zone soil moisture prognostic variables (*srfexc* and *rzexc*) can be twice as large as the typical magnitude of the respective monthly average increments. For *catdef*, by contrast, the intramonthly standard deviation of the increments tends to be much smaller than the typical magnitude of the increment. The large variability of the daily computed increments within a month (especially

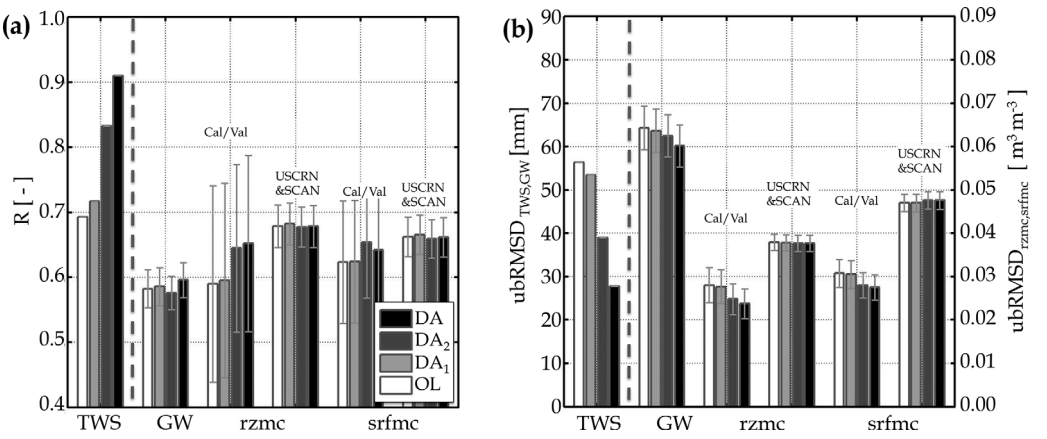

Figure 8. (a) Anomaly correlation coefficient (R) and (b) unbiased root-mean-squared difference (ubRMSD) for the open-loop (*OL*) and the GRACE-TWS assimilation schemes (DA_1, DA_2, *DA*, sections 2.3.3 and 2.3.4) when compared to independent in situ measurements of groundwater (GW), root-zone soil moisture (rzmc), and surface soil moisture (srfmc). Metrics for TWS are computed against the assimilated (scaled) GRACE-TWS observations. The mean values across the sites and the 95% confidence intervals are obtained after clustering of the sites. Soil moisture metrics for the sparse network sites are computed from the available sites in the Soil Climate Analysis Network (SCAN) and the U.S. Climate Reference Network (USCRN). Vertical dashed lines separate the TWS evaluation from the validation versus independent in situ measurements.

in root-zone and surface soil moisture) motivates the use of a monthly averaged increment in the data assimilation system, rather than subjectively choosing either the beginning of the month (as in DA_1) or the end of the month (as in DA_2).

3.3. Various Data Assimilation System Experiments

3.3.1. Calculation and Application of the Assimilation Increments

This section compares the three assimilation systems that are listed in Table 2. Recall that the assimilation systems differ by the way they calculate and apply the increments (section 2.3.4). Figure 8 reports the R and ubRMSD metrics obtained from averaging the metrics across individual validation sites (section 2.4.3).

The skill metrics for TWS verify the agreement between the assimilated observations and the model analyses over the entire CONUS domain as an internal check of the assimilation system. By design, the TWS metrics improve with assimilation. The TWS correlation statistics are 0.69 for *OL*, 0.72 for DA_1, 0.83 for DA_2, and 0.91 for *DA*. The ubRMSD values for TWS are 56 mm for *OL*, 53 mm for DA_1, 39 mm for DA_2, and 28 mm for *DA*. The *DA* scheme brings the monthly TWS analyses closest to the GRACE-TWS observations, because the increment is applied in its entirety at the beginning of the month and generally persists throughout the month. Applying the entire increment at the beginning of the month is the only way to ensure that the monthly average TWS output of the second model iteration is consistent with the TWS "analysis" that would be obtained directly from the update equation (equation (7)). If we applied only $1/N_{days}$ of the increment each day, we would only have caught up to the desired TWS at the end of the month, and for about half of the month (on average) we would still be closer to the forecasted TWS than to the "analysis" TWS. The practice of applying the update incrementally, as in method DA_1, was motivated by the desire to avoid shocks to the system and obtain a smoothly varying TWS time series from the assimilation. Our results show that this approach yields slightly worse skill metrics in the validation against independent observations.

For groundwater, the *OL* correlation coefficient equals 0.58 ± 0.03, where ± 0.03 describes the 95% confidence interval. The assimilation correlation coefficient values are 0.58 ± 0.03 for DA_1, 0.57 ± 0.03 DA_2, and 0.60 ± 0.03 for *DA*. The ubRMSD skill values are 64 ± 5 mm for *OL*, 64 ± 5 mm for DA_1, 63 ± 5 mm for DA_2, and 60 ± 5 mm for *DA*. The changes in the groundwater correlation and ubRMSD skill metrics compared to the *OL* metrics for any of the assimilation cases are not statistically significant. However, the best skill values are found for the *DA* case. Statistical significance is hard to obtain with monthly data (limited sample size), and a few isolated sites with degraded performance greatly impact the cluster-averaged statistics. Nevertheless, there are small but consistent improvements obtained with the proposed scheme (*DA*) over the model-only simulations (*OL*).

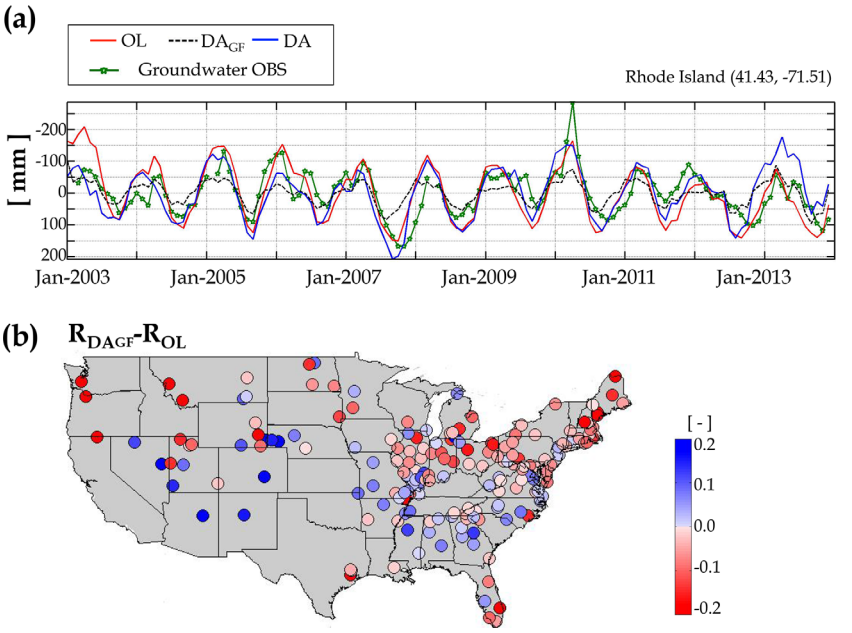

Figure 9. (a) Time series of catchment deficit for (red solid lines) *OL*, (black dashed lines) assimilation using TWS retrievals scaled by gain factors (Figure 1a) provided with the GRACE data product (DA_{GF}), and (blue solid lines) assimilation using TWS retrievals scaled with the ratio of CLSM (Figure 1b) and GRACE-TWS variabilities (*DA*). Green dots show in situ groundwater observations. (b) Difference in correlation skill between DA_{GF} and open-loop (*OL*; no assimilation) estimates for groundwater (i.e., same as Figure 5b but for DA_{GF} instead of *DA*).

For surface and root-zone soil moisture, the *DA* and DA_2 cases yield improvements in both R and ubRMSD at the Cal/Val sites, but improvements are again not statistically significant. For example, surface soil moisture R values are 0.62 for *OL*, 0.62 for DA_1, 0.65 for DA_2, and 0.64 for *DA*, with 95% confidence intervals of about $\pm$0.09. Root-zone soil moisture correlation skill values are 0.59 for *OL*, 0.59 for DA_1, 0.64 for DA_2, and 0.65 for *DA*, with 95% confidence intervals of about $\pm$0.14. Similarly, at the SCAN and USCRN sites the surface and root-zone soil moisture skill values are not statistically significantly different from those of the *OL* for any of the assimilation cases.

Given that GRACE-TWS assimilation has the most impact in the deeper layers (less intramonthly variability, section 3.2), the theoretical advantage of calculating the increments as a monthly average becomes marginal for the surface and root-zone soil moisture layers (higher intramonthly variability, section 3.2).

3.3.2. Effects of the Observation Scaling Parameters

The following illustrates how different scaling parameters (section 2.2) affect the data assimilation results. Figure 9a shows an example of catchment deficit time series for the *OL* and two data assimilation experiments, *DA* (sections 2.3.3 and 2.3.4) and DA_{GF}, which only differ in how the observations are scaled prior to assimilation. Recall that *DA* uses scaling parameters derived from the variability of CLSM output and the truncated/smoothed GRACE-TWS observations (equation (2)). Experiment DA_{GF} instead multiplies GRACE-TWS by the JPL-derived gain factors (Figure 1a) prior to assimilation. At the location shown in the figure, the scaling parameters are different and groundwater observations are available for reference. In experiment DA_{GF}, the amplitude of the observed TWS at this location is reduced prior to assimilation because the JPL gain factor equals 0.67. By contrast, the scaling parameters derived for the assimilation system for this location describe an amplification of the signal, with σ_x/σ_{y^*} equal to 2.67. As can be seen in Figure 9a, the use of the JPL gain factors provided with the GRACE-TWS observations (DA_{GF}) can result in obvious inconsistencies between the observed and modeled dynamic range of TWS.

The effects of the inconsistent dynamic range impact the skill values in the groundwater validation. The differences in groundwater correlation skills between the DA_{GF} case and the open-loop are shown in Figure 9b. By comparing this result to that of the nominal *DA* case (Figure 5b), it becomes clear that the JPL-provided gain factors are not suitable for use in our data assimilation system. Most of the differences between the *DA* and DA_{GF} improvements are seen where the scaling parameters (i.e., Figures 1a and 1b) differ the most. This is the case, for example, in the North East and along the Mississippi River. Groundwater

Table 4. Mean, Median (Q50), and the Interquantile Range (Q25, Q75) of the Correlation Coefficient (R) and the Unbiased Root-Mean-Square Difference (ubRMSD) Across All Groundwater Validation Locations for Estimates From the Data Assimilation Scheme With JPL-Derived Scaling Parameters (DA_{GF}, Section 3.3.2)[a]

			R		ubRMSD		
	N. sites		Mean (*CI*)	Q50 (Q25, Q75)	Mean (*CI*)	Q50 (Q25, Q75)	
GW	181	*OL*	0.58 (±0.03)	0.60 (0.51, 0.71)	64 (±5)	62 (51, 75)	mm
		DA	0.60 (±0.03)	0.64 (0.53, 0.74)	60 (±5)	56 (45, 69)	mm
		DA_{GF}	0.57 (±0.03)	0.58 (0.45, 0.70)	60 (±5)	57 (45, 71)	mm
		% improved	37%		72%		

[a]The table also reports statistics for the open-loop (*OL*) and the data assimilation (*DA*) cases (Table 3). Mean statistics and 95% confidence intervals (*CI*) are obtained from clustering of the sites (section 2.4).

bulk statistic skills for the experiment DA_{GF} are reported in Table 4. While the ubRMSD for DA_{GF} and *DA* both equal 60 ± 5 mm, the correlation skill for DA_{GF} is equal to 0.57 ± 0.03, which is worse than the corresponding value for the *DA* case (0.60 ± 0.03) and the open-loop case (0.58 ± 0.03). Overall, the improvements due to GRACE-TWS data assimilation in experiment DA_{GF} are worse than those obtained from experiment *DA*. It is therefore important that TWS observations are scaled using scaling parameters that ensure a climatologically consistent assimilation system.

4. Summary and Conclusions

Because of the unique spatial and temporal resolution of GRACE-TWS observations, it is not obvious how best to assimilate such data into a land surface model. This work revisits various assimilation approaches and proposes an alternative algorithm to integrate gridded GRACE-TWS observations within a land surface model with the objective of improving groundwater and soil moisture estimates. Special attention is paid to (i) the calculation and application of the increments and (ii) the careful design of a climatologically consistent assimilation system.

The main findings of the presented work can be summarized as follows.

1. The assimilation system partitions the vertically integrated GRACE-TWS column of water into the various water storage compartments (i.e., surface and root-zone soil moisture, groundwater, and snow), in accordance with prior model information of their relative contribution and uncertainties to TWS. The assimilation of GRACE-TWS primarily affects (in absolute terms) deeper moisture storages (i.e., groundwater), whereas the impact on root-zone and surface layer soil moisture is smaller. These results motivate future efforts to combine GRACE-TWS observations with observations that are more sensitive to surface soil moisture such as observations from the SMOS or SMAP missions for a more comprehensive improvement of the entire soil water profile.
2. The large variability of the daily computed increments within a month (especially in root-zone and surface soil moisture) motivates the use of a monthly averaged increment in a GRACE-TWS data assimilation system, rather than computing the increment for just a single day at the beginning or at the end of the month as in existing GRACE assimilation schemes. This theoretically more attractive approach yields a small benefit in monthly groundwater and soil moisture estimation skill compared with the existing assimilation methods.
3. The assimilation of GRACE-TWS is affected by the use of observation scaling parameters. Multiplicative gain factors are provided with the GRACE data product. These gain factors are essential for data analysis because they restore the signal lost during the truncation and smoothing needed to retrieve the GRACE-TWS observations. However, the factors provided with the product are not necessarily useful in a data assimilation system. Such assimilation systems expect observations with similar long-term properties as the land surface simulations, and they only aim at correcting for nonsystematic (short-term) errors. To ensure that the assimilation system is not adversely affected by systematic differences between the model and the TWS observations, the model would ideally be recalibrated so that its TWS climatology matches that of the observations. If this is not possible, it is recommended to ensure climatological consistency between observations and simulations prior to data assimilation through scaling using model-specific parameters.

Acknowledgments
The authors thank Mike Cosh and Tom Jackson for providing the in situ data for the SMAP core validation watersheds, Qing Liu for her help with data quality control, Randy Koster for his help with CLSM, and Bailing Li and Rasmus Houborg for their help with the groundwater data. GRACE-TWS was received from "http://GRACE.jpl.nasa.gov," which used the "Physical Oceanography Distributed Active Archive Center." Computational resources were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at the Goddard Space Flight Center. This study was supported by the NASA Terrestrial Hydrology program.

References

Brodzik, M. J., B. Billingsley, T. Haran, B. Raup, and M. H. Savoie (2012), Ease-grid 2.0: Incremental but significant improvements for earth-gridded data sets, *ISPRS Int. J. Geo-Inf.*, *1*(1), 32–45.
Burgers, G., P. Jan van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*(6), 1719–1724.
Cosh, M. H., T. J. Jackson, S. Moran, and R. Bindlish (2008), Temporal persistence and stability of surface soil moisture in a semi-arid watershed, *Remote Sens. Environ.*, *112*(2), 304–313.
De Lannoy, G. J., and R. H. Reichle (2015), Global assimilation of multi-angle and multi-polarization smos brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation, *J. Hydrometeorol.*, *17*, 669–691.
De Lannoy, G. J., R. H. Reichle, P. R. Houser, V. Pauwels, and N. E. Verhoest (2007), Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, *Water Resour. Res.*, *43*, W09410, doi:10.1029/2006WR005449.
De Lannoy, G. J., R. H. Reichle, P. R. Houser, K. R. Arsenault, N. E. Verhoest, and V. R. Pauwels (2010), Satellite-scale snow water equivalent assimilation into a high-resolution land surface model, *J. Hydrometeorol.*, *11*(2), 352–369.
De Lannoy, G. J., R. D. Koster, R. H. Reichle, S. P. Mahanama, and Q. Liu (2014), An updated treatment of soil texture and associated hydraulic properties in a global land modeling system, *J. Adv. Model. Earth Syst.*, *6*, 957–979, doi:10.1002/2014MS000330.
Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Q. J. R. Meteorol. Soc.*, *131*(613), 3385–3396.
Diamond, H. J., et al. (2013), US Climate Reference Network after one decade of operations: Status and assessment, *Bull. Am. Meteorol. Soc.*, *94*(4), 485–498.
Draper, C., R. Reichle, G. De Lannoy, and B. Scarino (2015), A dynamic approach to addressing observation-minus-forecast bias in a land surface skin temperature data assimilation system, *J. Hydrometeorol.*, *16*(1), 449–464.
Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration, *J. Geophys. Res.*, *105*(D20), 24,823–24,838.
Eicker, A., M. Schumacher, J. Kusche, P. Döll, and H. M. Schmied (2014), Calibration/data assimilation approach for integrating GRACE data into the watergap global hydrology model (wghm) using an ensemble Kalman filter: First results, *Surv. Geophys.*, *35*(6), 1285–1309.
Entekhabi, D. (2014), *Coauthors (2014), SMAP Handbook, JPL Publ., 400*, 1567 pp.
Entekhabi, D., et al. (2010), The soil moisture active passive (SMAP) mission, *Proc. IEEE*, *98*(5), 704–716.
Famiglietti, J. S., and M. Rodell (2013), Water in the balance, *Science*, *340*(6138), 1300–1301.
Forman, B. A., and R. Reichle (2013), The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system, *Water Resour. Res.*, *49*, 7457–7468, doi:10.1002/2012WR012885.
Forman, B. A., R. Reichle, and M. Rodell (2012), Assimilation of terrestrial water storage from GRACE in a snow-dominated basin, *Water Resour. Res.*, *48*, W01507, doi:10.1029/2011WR011239.
Frappart, F., G. Ramillien, S. Biancamaria, N. M. Mognard, and A. Cazenave (2006), Evolution of high-latitude snow mass derived from the GRACE gravimetry mission (2002–2004), *Geophys. Res. Lett.*, *33*, L02501, doi:10.1029/2005GL024778.
Hirschi, M., B. Mueller, W. Dorigo, and S. Seneviratne (2014), Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability, *Remote Sens. Environ.*, *154*, 246–252.
Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik (2012), Drought indicators based on model-assimilated gravity recovery and climate experiment (GRACE) terrestrial water storage observations, *Water Resour. Res.*, *48*, W07525, doi:10.1029/2011WR011291.
Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. Seyfried, D. C. Goodrich, M. S. Moran, and J. Du (2010), Validation of advanced microwave scanning radiometer soil moisture products, *IEEE Trans. Geosci. Remote Sens.*, *48*(12), 4256–4272.
Jackson, T. J., et al. (2012), Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the US, *IEEE Trans. Geosci. Remote Sens.*, *50*(5), 1530–1543.
Konikow, L. F. (2013), *Groundwater Depletion in the United States (1900–2008)*, U.S. Dep. of the Inter., U.S. Geol. Surv., John Wiley & Sons, Ltd., USA.
Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, *105*(D20), 24,809–24,822.
Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma (2009), On the nature of soil moisture in land surface models, *J. Clim.*, *22*(16), 4322–4335.
Koster, R. D., et al. (2010), Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment, *Geophys. Res. Lett.*, *37*, L02402, doi:10.1029/2009GL041677.
Kumar, S. V., et al. (2016), Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System, *J. Hydrometeorol.*, doi:10.1175/JHM-D-15-0157.1, in press.
Lahoz, W., B. Khattatov, and R. Ménard (2010), Data assimilation and information, in *Data Assimilation*, pp. 3–12, Springer Berlin Heidelberg.
Landerer, F., and S. Swenson (2012), Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, *48*, W04531, doi:10.1029/2011WR011453.
Li, B., and M. Rodell (2014), Evaluation of a model-based groundwater drought indicator in the conterminous US, *J. Hydrol.*, *526*, 78–88.
Li, B., M. Rodell, B. F. Zaitchik, R. H. Reichle, R. D. Koster, and T. M. van Dam (2012), Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe, *J. Hydrol.*, *446*, 103–115.
Liu, Q., R. H. Reichle, R. Bindlish, M. H. Cosh, W. T. Crow, R. de Jeu, G. J. De Lannoy, G. J. Huffman, and T. J. Jackson (2011), The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system, *J. Hydrometeorol.*, *12*(5), 750–765.
Long, D., L. Longuevergne, and B. R. Scanlon (2015), Global analysis of approaches for deriving total water storage changes from GRACE satellites, *Water Resour. Res.*, *51*, 2574–2594, doi:10.1002/2014WR016853.
Luthcke, S. B., T. Sabaka, B. Loomis, A. Arendt, J. McCarthy, and J. Camp (2013), Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution, *J. Glaciol.*, *59*(216), 613–631.
Niu, G.-Y., K.-W. Seo, Z.-L. Yang, C. Wilson, H. Su, J. Chen, and M. Rodell (2007), Retrieving snow mass from GRACE terrestrial water storage change with a land surface model, *Geophys. Res. Lett.*, *34*, L15704, doi:10.1029/2007GL030413.
Reager, J., and J. Famiglietti (2009), Global terrestrial water storage capacity and flood potential using GRACE, *Geophys. Res. Lett.*, *36*, L23402, doi:10.1029/2009GL040826.
Reager, J., B. Thomas, and J. Famiglietti (2014), River basin flood potential inferred using GRACE gravity observations at several months lead time, *Nat. Geosci.*, *7*(8), 588–592.

Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, *4*(6), 1229–1242.
Reichle, R. H., G. J. De Lannoy, B. A. Forman, C. S. Draper, and Q. Liu (2014), Connecting satellite observations with water cycle variables through land data assimilation: Examples using the NASA GEOS-5 LDAS, in *The Earth's Hydrological Cycle*, pp. 577–606, Springer Netherlands.
Rienecker, M. M., et al. (2011), MERRA: NASA's modern-era retrospective analysis for research and applications, *J. Clim.*, *24*(14), 3624–3648.
Rodell, M., J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson (2007), Estimating groundwater storage changes in the Mississippi river basin (USA) using GRACE, *Hydrogeol. J.*, *15*(1), 159–166.
Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature*, *460*(7258), 999–1002.
Rodell, M., E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoing, and J. Nigro (2011), Estimating evapotranspiration using an observation based terrestrial water budget, *Hydrol. Processes*, *25*(26), 4082–4092.
Rowlands, D., S. Luthcke, S. Klosko, F. Lemoine, D. Chinn, J. McCarthy, C. Cox, and O. Anderson (2005), Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements, *Geophys. Res. Lett.*, *32*, L04310, doi:10.1029/2004GL021908.
Schaefer, G. L., M. H. Cosh, and T. J. Jackson (2007), The USDA natural resources conservation service soil climate analysis network (SCAN), *J. Atmos. Oceanic Technol.*, *24*(12), 2073–2077.
Stieglitz, M., A. Ducharne, R. Koster, and M. Suarez (2001), The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales, *J. Hydrometeorol.*, *2*(3), 228–242.
Su, H., Z.-L. Yang, R. E. Dickinson, C. R. Wilson, and G.-Y. Niu (2010), Multisensor snow data assimilation at the continental scale: The value of gravity recovery and climate experiment terrestrial water storage information, *J. Geophys. Res.*, *115*, D10104, doi:10.1029/2009JD013035.
Swenson, S., and J. Wahr (2002), Methods for inferring regional surface-mass anomalies from gravity recovery and climate experiment (GRACE) measurements of time-variable gravity, *J. Geophys. Res.*, *107*(B9), 2193, doi:10.1029/2001JB000576.
Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, *33*, L08402, doi:10.1029/2005GL025285.
Swenson, S., P. J.-F. Yeh, J. Wahr, and J. Famiglietti (2006), A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois, *Geophys. Res. Lett.*, *33*, L16401, doi:10.1029/2006GL026962.
Tangdamrongsub, N., S. Dunne, B. C. Gunter, P. Ditmar, and A. Weerts (2014), Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine river basin, *Hydrol. Earth Syst. Sci. Discuss.*, *19*, 2079–2100.
Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004), GRACE measurements of mass variability in the earth system, *Science*, *305*(5683), 503–505.
Thomas, A. C., J. T. Reager, J. S. Famiglietti, and M. Rodell (2014), A GRACE-based water storage deficit approach for hydrological drought characterization, *Geophys. Res. Lett.*, *41*, 1537–1545, doi:10.1002/2014GL059323.
van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning (2014), A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, *Hydrol. Earth Syst. Sci.*, *18*, 2955–2973.
Velicogna, I., T. Sutterley, and M. van den Broeke (2014), Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data, *Geophys. Res. Lett.*, *41*, 8130–8137, doi:10.1002/2014GL061052.
Voss, K. A., J. S. Famiglietti, M. Lo, C. Linage, M. Rodell, and S. C. Swenson (2013), Groundwater depletion in the middle east from GRACE with implications for transboundary water management in the Tigris-Euphrates-western Iran region, *Water Resour. Res.*, *49*, 904–914, doi:10.1002/wrcr.20078.
Wahr, J., S. Swenson, and I. Velicogna (2006), Accuracy of GRACE mass estimates, *Geophys. Res. Lett.*, *33*, L06401, doi:10.1029/2005GL025305.
Winter T. C., J. W. Harvey, Q. L. Franke, and W. M. Alley 1998), Ground water and surface water: A single resource, *U.S. Geol. Surv. Circ.*, *1139*, 79 pp.
Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008), Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi river basin, *J. Hydrometeorol.*, *9*(3), 535–548.